



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/338,176	06/22/1999	HEUNG-YEUNG SHUM	3382-52053	1062
26119	7590	03/13/2006	EXAMINER	
KLARQUIST SPARKMAN LLP 121 S.W. SALMON STREET SUITE 1600 PORTLAND, OR 97204				WONG, ALLEN C
ART UNIT		PAPER NUMBER		
		2613		

DATE MAILED: 03/13/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>
	09/338,176	SHUM ET AL.
	<b>Examiner</b>	<b>Art Unit</b>
	Allen Wong	2613

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) Responsive to communication(s) filed on 02 December 2005.
- 2a) This action is **FINAL**.      2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) Claim(s) 1,2 and 4-37 is/are pending in the application.
  - 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) Claim(s) \_\_\_\_\_ is/are allowed.
- 6) Claim(s) 1,2 and 4-37 is/are rejected.
- 7) Claim(s) \_\_\_\_\_ is/are objected to.
- 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on \_\_\_\_\_ is/are: a) accepted or b) objected to by the Examiner.
 

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a) All    b) Some \* c) None of:
    1. Certified copies of the priority documents have been received.
    2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

#### Attachment(s)

- |  |   |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)                     |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                     | Paper No(s)/Mail Date. _____ .  |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ . | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
|  | 6) <input type="checkbox"/> Other: _____ .                                  |

## DETAILED ACTION

### ***Response to Arguments***

1. Applicant's arguments filed 12/2/05 have been fully read and considered but they are not persuasive.

Regarding lines 10-11, 15-16 and 18-20 on page 13 of applicant's remarks about claim 23, applicant asserts that Lee does not teach the limitation of "adding" based on "if a threshold number of feature points are identified in the next frame, adding the next frame, adding the next frame to the first segment. The examiner respectfully disagrees.

Jain teaches manually adjusting the number of key frames, where the number is one key frame for every thirty frames, ie. a segment (col.23, ln.64 to col.24, ln.3). Therefore, since Jain teaches manually adjusting one key frame or representative frame for every thirty frames, it would have been obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene. Also, in column 2, line 65 to column 3, line 31, Lee teaches the determining whether a threshold number of feature points from base frame are identified in the second frame by using threshold values TH and comparison of threshold values of feature points between the current frame and the reference frame to check if the threshold is exceeded, and that Lee also teaches that if a threshold number of feature points are identified in the second frame, adding the second frame to the segment. In figure 3, Lee suggests the cyclical process of determination of the threshold number values.

Thus, Lee teaches the repeating the analyzing step, determining step and adding step for subsequent frames until the number of feature points in a frame falls below the threshold number.

Therefore, it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole, for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner, as disclosed in Lee's column 2, lines 60-64.

Dependent claims 24-30 are rejected for at least similar reasons as stated for claim 23.

Dependent claims 17 and 20 are rejected for at least similar rationale as stated for claim 23.

Independent claim 1 is now rejected under 35 U.S.C. 103 in view of Jain and Lee since claim 1 now comprises some similar limitations as mentioned in claim 23. See above paragraphs and the rejection below.

Dependent claims 2 and 4-7 are rejected for at least similar reasons as claim 1. For instance, in column 23, line 58 to column 24, line 3, Jain discloses the use of extracting of key frames by selecting one key frame from every 30 frames. Also peruse the rejection below.

Regarding lines 11-17 on page 19 of applicant's remarks about claim 9, applicant argues that Jain does not disclose "... for each segment, encoding the frames in the segment into at least two virtual key frames that include a three-dimensional structure

for the segment and an uncertainty associated with the segment". The examiner respectfully disagrees. In column 23, line 58 to col.24, line 3, Jain discloses extracting virtual key frames by choosing one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames. Also, in column 24, lines 38-67, Jain discloses using virtual key frames for obtaining the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from virtual key frames. Thus, the segmented frames are encoded into at least two virtual key frames to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization. In fig.8, camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second at the standard NTSC frame rate (30 frames/sec), where the virtual key frames are extracted in preparation for image recording and encoding sent for processing to be viewed at the display terminal. Thus, Jain meets the broad limitations of claim 9.

Dependent claims 10-22 are rejected for at least similar reasons as claim 9.

Regarding lines 2-4 on page 21 of applicant's remarks about claim 36, applicant contends that Jain does not specifically disclose "calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames" and "extracting virtual key frames from each partial model". The examiner respectfully disagrees. In figure 12, Jain discloses that there are multiple "image to ground projection" sections that are used for calculating and projecting an image or a partial model for each segment of that includes three-dimensional

occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model. In column 21, line 63 to column 22, line 7, Jain discloses the obtaining of the feature points within the frames. And in column 22, line 62 to column 23, line 56, Jain discloses the equations including three dimensional coordinates (x, y, z) along with camera position or pose, camera angle and camera parameter to obtain a partial model or a "image to ground projection". Thus, Jain discloses "calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames".

In column 23, line 58 to column 24, line 3, Jain teaches extracting key frames by selecting one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames. And also, in column 24, line 38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization. Thus, Jain discloses "extracting virtual key frames from each partial model".

Regarding lines 1-5 on page 23 of applicant's remarks about claim 37, applicant argues that Jain does not disclose "calculating a partial model for each segment..." and "extracting virtual key frames from each partial model". The examiner respectfully disagrees. Claim 37 is rejected for at least similar reasons as claim 36.

Regarding lines 2-4 and 10-14 on page 24 of applicant's remarks about claim 31, applicant mentions that Jain does not explicitly suggest "dividing a long sequence of frames into segments and reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment", and that "manual selection" is not the same as "manual adjustment". The examiner respectfully disagrees. In figure 8, Jain discloses camera 1 captures a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments. Also, Jain states that the extraction of key frames by selecting one key frame from every 30 frames (ie. a segment of a sequence of frames), as disclosed in column 23, line 58 to column 24, line 3. Clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames.

Also, in column 23, line 64 to column 24, line 3, Jain discloses manually adjusting the number of key frames, where the number (segment) is one key frame for every thirty frames. Therefore, since Jain teaches the manual adjustment of one key frame or representative frame for every thirty frames, it would have been reasonably obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation

of the targeted scene. Furthermore, the term “manual adjustment” can imply many things and that “manual selection” can be interpreted as one form of “manual adjustment” since selection is a type of adjustment.

Dependent claims 32-35 are rejected for at least similar reasons as claim 31.

Thus, the rejection is maintained.

Regarding applicant’s request for interview, the applicant is invited to telephonically request an interview if there be any issues with regard to this Office Action that applicant may feel deemed necessary to further support their position in addition to the amendment filed 12/2/05. Otherwise, a written after final response should be fully adequate to state the applicant’s position on this case.

### ***Claim Rejections - 35 USC § 102***

1. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

2. Claims 9-16, 18, 19, 21, 22, 36 and 37 are rejected under 35 U.S.C. 102(b) as being anticipated by Jain et al (5,729,471).

Regarding claims 9 and 21, discloses a method of recovering a three-dimensional scene from two-dimensional images, the method comprising:

identifying a sequence of two-dimensional frames that include two-dimensional images (fig.12, note camera 1 obtain video images in two-dimensional form; also see

fig.8, note camera 1 obtains a sequence of two-dimensional images, and cameras 2 and 3 also obtain a corresponding sequence of images; col.22, ln.1-3);

dividing the sequence of images into segments, wherein a segment includes a plurality of frames (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments; also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames);

for each segment, encoding the frames in the segment into at least two virtual frames that include a three-dimensional structure for the segment and an uncertainty associated with the segment (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of virtual key frames by selecting one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames; also, col.24, ln.38-67, Jain discloses the virtual key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from virtual key frames, thus, segmented frames are encoded into at least two virtual key frames to ascertain the best, possible three-

dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 10, Jain discloses the identifying the base frame, identifying the feature points in the base frame, and defining the segments (col.21, ln.63 to col.22, ln.7; Jain discloses the identification of feature points in the plural frames that includes the first base frame in the segments from the sequence of images).

Regarding claim 11, Jain discloses the variation of segments and variation of frames (fig.8, note camera 1 has multiple 413 frames in approximately 13 seconds, where each segment has 30 frames to obtain approximately 13 segments from camera 1, whereas camera 2 has 181 frames in 6 seconds, or approximately 6 segments from camera 2, etc.).

Regarding claim 12, Jain discloses identify feature points, estimating three dimensional coordinates, and estimating camera rotation and translation (fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a “image to ground projection”).

Regarding claims 13-16, Jain discloses the performance of a two-frame structure from motion algorithm on each of the segments to create a partial model (fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a “image to ground projection”); and eliminating ambiguity (col.24, ln.38-67, Jain discloses the virtual key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from virtual key frames, thus, segmented frames are encoded into at least two virtual key frames to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 18, Jain discloses extracting virtual key frames (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction

of the two-dimensional frame data to yield the 3D visualization) and bundle adjustment of key frames (fig.12, note the “3D visualization” section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 19, Jain discloses performing motion estimation to identify feature points (col.21, ln.63 to col.22, ln.7).

Regarding claim 22, Jain discloses the use of a computer-readable medium to execute instructions for performing the method of claim 9 (col.15, ln.65-67).

Regarding claim 36, Jain discloses a computer-readable medium having computer-executable instructions for performing a method comprising:

providing a sequence of two-dimensional frames (fig.12, note camera 1 obtain video images in two-dimensional form; also see fig.8, note camera 1 obtains a sequence of two-dimensional images, and cameras 2 and 3 also obtain a corresponding sequence of images; col.22, ln.1-3);

dividing the sequence into segments (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be

considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments; also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames);

calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames (fig.12, note there are multiple "image to ground projection" sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a "image to ground projection");

extracting virtual key frames from each partial model, the virtual key frames having three-dimensional coordinates for the frames and an uncertainty associated with the frames (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames; also, col.24, ln.38-67, Jain discloses

the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization); and

bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames (fig.12, note the “3D visualization” section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 37, Jain discloses an apparatus for recovering a three-dimensional scene from a sequence of two-dimensional frames by segmenting the frames, comprising:

means for capturing two-dimensional images (fig.12, note camera 1 obtain video images in two-dimensional form; also see fig.8, note camera 1 obtains a sequence of two-dimensional images, and cameras 2 and 3 also obtain a corresponding sequence of images; col.22, ln.1-3);

means for dividing the sequence into segments (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments; also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames);

means for calculating a partial model for each segment that includes three-dimensional coordinates and camera pose for features within the frames (fig.12, note there are multiple "image to ground projection" sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a "image to ground projection");

means for extracting virtual key frames from each partial model (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from

every 30 frames in that every 30 frames can be considered a segment of a sequence of frames); and

means for bundle adjusting the virtual key frames to obtain a complete three-dimensional reconstruction of the two-dimensional frames (fig.12, note the “3D visualization” section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

### ***Claim Rejections - 35 USC § 103***

3. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

4. Claims 1-2, 4-8, 17, 20 and 23-30 are rejected under 35 U.S.C. 103(a) as being unpatentable over Jain et al (5,729,471) in view of Lee (5,612,743).

Regarding claim 1, Jain discloses a method of recovering a three-dimensional scene from two-dimensional images, the method comprising:

providing a sequence of frames (fig.12, note camera 1 obtain video images in two-dimensional form; also see fig.8, note camera 1 obtains a sequence of two-dimensional images, and cameras 2 and 3 also obtain a corresponding sequence of images; col.22, ln.1-3);

dividing the sequence of frames into frame segments wherein the frames in the sequence comprise feature points and wherein dividing the sequence of frames into frame segments (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a frame segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments; also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames; also fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that

includes camera position or pose, camera angle and camera parameter to obtain a partial model or a “image to ground projection”);

performing three-dimensional reconstruction individually for each frame segment derived by dividing the sequence of frames (fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a “image to ground projection”); and

combining the three-dimensional reconstructed segments together to recover a three-dimensional scene for the sequence of images (fig.12, note the “3D visualization” section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Jain does not specifically disclose the determining whether a threshold number of feature points being tracked between the frames of the frame segments. However, Lee teaches the determining whether a threshold number of feature points being tracked between the frames of the frame segments (col.2, ln.65 to col.3, ln.31; Lee teaches the use of threshold values TH and comparison of threshold values of feature points between the current frame and the reference frame to check if the threshold is exceeded). Therefore, it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole, for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner (col.2, ln.60-64).

Regarding claim 2, Jain discloses the use of virtual key frames (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames).

Regarding claim 4, Jain discloses the performance of a two-frame structure from motion algorithm on each of the segments to create a partial model (fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera

Art Unit: 2613

position or pose, camera angle and camera parameter to obtain a partial model or a "image to ground projection"); and eliminating ambiguity (col.24, ln.38-67, Jain discloses the virtual key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from virtual key frames, thus, segmented frames are encoded into at least two virtual key frames to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claims 5 and 7, Jain discloses extracting virtual key frames (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization) and bundle adjustment of key frames (fig.12, note the "3D visualization" section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-

dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 6, Jain discloses identifying feature points, estimating three dimensional coordinates, and estimating camera rotation and translation (fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a “image to ground projection”).

Regarding claim 8, Jain discloses the use of a computer-readable medium to execute instructions for performing the method of claim 1 (col.15, ln.65-67).

Regarding claims 17, 23, 24 and 28, Jain discloses a method of recovering a three-dimensional scene from a sequence of two-dimensional frames, comprising:

identifying at least a first base frame in a sequence of two dimensional frames (fig.12, note camera 1 obtain video images in two-dimensional form; also see fig.8, note camera 1 obtains a sequence of two-dimensional images, and cameras 2 and 3 also obtain a corresponding sequence of images; see col.22, ln.1-3; fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec),

can be considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments; also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly, Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames);

adding the at least first base frame to create a first segment of the sequence (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a frame segment, so in this case, camera 1 has approximately 14 frame segments, so a first segment of the sequence is created);

identifying feature points in at least a first base frame in a first segment (col.21, ln.63 to col.22, ln.7; Jain discloses the identification of feature points in the plural frames that includes the first base frame); and

analyzing a second frame in the segment to identify the feature points in the second frame (col.21, ln.63 to col.22, ln.7; Jain discloses the identification of feature points in each frame from a plurality of frames that includes the second frame).

Jain does not specifically disclose the adding the second frame to the segment. However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, ie. a segment (col.23, ln.64 to col.24,

In.3). Therefore, since Jain teaches the manual adjustment of one key frame or representative frame for every thirty frames, it would have been obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene.

Jain does not specifically disclose the determining whether a threshold number of feature points from base frame are identified in the second frame; if a threshold number of feature points are identified in the second frame, adding the second frame to the segment; and repeating the analyzing step, determining step and adding step for subsequent frames until the number of feature points in a frame falls below the threshold number. However, Lee teaches the determining whether a threshold number of feature points from base frame are identified in the second frame (col.2, ln.65 to col.3, ln.31; Lee teaches the use of threshold values TH and comparison of threshold values of feature points between the current frame and the reference frame to check if the threshold is exceeded); if a threshold number of feature points are identified in the second frame, adding the second frame to the segment (col.2, ln.65 to col.3, ln.31); and repeating the analyzing step, determining step and adding step for subsequent frames until the number of feature points in a frame falls below the threshold number (fig.3, note Lee discloses the process is cyclical and repetitive, thus the analysis, determination and addition steps are repeated). Therefore, it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole,

for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner (col.2, ln.60-64).

Regarding claim 20, Jain does not specifically disclose creating a template block in a first frame, creating a search window used in the second frame, and comparing an intensity difference between the search window and the template block to locate the feature point in the second frame. However, Lee teaches that creating a template block in a first frame, creating a search window used in the second frame, and comparing an intensity difference between the search window and the template block to locate the feature point in the second frame (fig.4, note frame A and frame B are the first and second frames, note fig.3, element 313 also discloses the comparison process to compare differences to determine or locate the feature point in the second frame). Therefore, it would have been obvious to one of ordinary skill in the art to combine the teachings of Jain and Lee, as a whole, for improving the encoding of video image data so as to accurately encode images via the selection of feature points according to the motion of objects in a financially robust manner (col.2, ln.60-64).

Regarding claim 25, Jain discloses performing motion estimation to identify feature points (col.21, ln.63 to col.22, ln.7).

Regarding claim 26, Jain discloses the identification of corners as feature points (col.22, ln.15-22; note the disclosure of borders, hashlines, marks are feature points to create corners as to determine camera status and pose).

Regarding claim 27, Jain discloses the number of frames can vary between segments (col.23, ln.64 to col.24, ln.3).

Regarding claim 29, Jain discloses the bundle adjustment of key frames (fig.12, note the “3D visualization” section is the product of the adjusting of the virtual key frames to produce a complete three-dimensional reconstruction of the two dimensional frames obtained by video camera 1 to video camera N; also, col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 30, Jain discloses the use of a computer-readable medium to execute instructions for performing the method of claim 23 (col.15, ln.65-67).

Claim 31-35 are rejected under 35 U.S.C. 103(a) as being unpatentable over Jain et al (5,729,471).

Regarding claim 31, Jain discloses a method of recovering a three-dimensional scene from a sequence of two-dimensional frames (fig.12), an improvement comprising: dividing a long sequence of frames into segments (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments;

also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly; Jain discloses there are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames),

wherein the representative frames are used to recover the three-dimensional scene and remaining frames are discarded so that three-dimensional scene is effectively compressed (col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of virtual key frames by selecting one key frame from every 30 frames in that every 30 frames can be considered a segment of a sequence of frames; also, col.24, ln.38-67, Jain discloses the virtual key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from virtual key frames, thus, segmented frames are encoded into at least two virtual key frames to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization, the excess remaining frames are discarded).

Jain does not specifically disclose the reducing the number of frames in each segment by representing the segments using between two and five representative frames per segment. However, Jain discloses the manual adjustment of the number of key frames, where the number is one key frame for every thirty frames, ie. a segment (col.23, ln.64 to col.24, ln.3). Therefore, since Jain teaches the manual adjustment of one key frame or representative frame for every thirty frames, it would have been

obvious to one of ordinary skill in the art to manually change the number of key (representative) frames per segment from anywhere between two to five key or representative frames per segment if necessary for accurately enhancing the three-dimensional representation of the targeted scene.

Regarding claim 32, Jain discloses that each representative frame have an associated uncertainty (col.24, ln.38-67, Jain discloses the key frames are used to obtain the best possible three-dimensional reconstruction of the two-dimensional frame data in that if there is not enough known points, ie. uncertainty, from key frames, estimates or bundle adjustments were made to ascertain the best, possible three-dimensional reconstruction of the two-dimensional frame data to yield the 3D visualization).

Regarding claim 33, Jain discloses the long sequence of frames includes over 75 frames (fig.8, note that camera 1 obtains a sequence of 412 frames, which clearly is over 75 frames).

Regarding claim 34, Jain discloses the division of the long sequence into segments and tracking feature points (fig.8, note that camera 1 obtains a sequence of 412 frames for approximately 13 seconds, and that every 30 frames obtained for each second, ie. the standard NTSC frame rate (30 frames/sec), can be considered a segment, so in this case, camera 1 has approximately 14 segments, thus, Jain discloses the division of the sequence of images into segments; also, in col.23, ln.58 to col.24, ln.3; Jain discloses the extraction of key frames by selecting one key frame from every 30 frames, ie. a segment of a sequence of frames, clearly, Jain discloses there

are segments within a sequence of frames, otherwise, the ascertainment of key frames would not be possible without these segments, where each segment is formed from a sequence of 30 frames; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames).

Regarding claim 35, Jain discloses the performance of a two-frame structure from motion algorithm on each of the segments to create a partial model (fig.12, note there are multiple “image to ground projection” sections that are used to calculate and project an image or a partial model for each segment of that includes three-dimensional occupancy estimation for which a 3D map of is generated in an attempt to form a dynamic model; col.21, ln.63 to col.22, ln.7, Jain discloses the obtaining of the feature points within the frames; col.22, ln.62 to col.23, ln.56, Jain discloses the use of equations that includes three dimensional coordinates (x, y, z) that includes camera position or pose, camera angle and camera parameter to obtain a partial model or a “image to ground projection”).

### ***Conclusion***

5. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any

Art Unit: 2613

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

***Contact Information***

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Allen Wong whose telephone number is (571) 272-7341. The examiner can normally be reached on Mondays to Thursdays from 8am-6pm Flextime.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, James J. Groody can be reached on (571) 272-7418. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Allen Wong  
Primary Examiner  
Art Unit 2613

AW